ADAPTIVE TARGET-SCALE-INVARIANT HYPERSPECTRAL ANOMALY DETECTION

João M. Romano
U.S. Army's Armament Research & Development Center (ARDEC)
Picatinny, New Jersey, 07806
joao.m.romano@pica.army.mil

Dalton Rosario
U.S. Army Research Laboratory (ARL)
Adelphi, Maryland
rosario@arl.army.mil

ABSTRACT
Ground to ground, sensor to object viewing perspective presents a major challenge for autonomous window based object detection, since object scales at this viewing perspective cannot be approximated. In this paper, we present a fully autonomous parallel approach to address this challenge. Using hyperspectral (HS) imagery as input, the approach features a random sampling stage, which does not require secondary information (range) about the targets; a parallel process is introduced to mitigate the inclusion by chance of target samples into clutter background classes during random sampling; and a fusion of results. The probability of sampling targets by chance within the parallel processes is modeled by the binomial distribution family, which can assist on tradeoff decisions. Since this approach relies on the effectiveness of its core algorithmic detection technique, we also propose a compact test statistic for anomaly detection, which is based on a principle of indirect comparison. This detection technique has shown to preserve meaningful detections (genuine anomalies in the scene) while significantly reducing the number of false positives (e.g. transitions of background regions). To capture the influence of parametric changes using both the binomial distribution family and actual HS imagery, we conducted a series of rigid statistical experiments and present the results in this paper.

KEY WORDS
Parallel anomaly detection, hyperspectral imagery, and random sampling

1. Introduction

Effective anomaly detection has a high value for applications where a priori information is unavailable about the objects of interest. Its value may be attributed to the simplicity built into this family of detection algorithms, where detectors search for *rare* pixels in the image whose information significantly differs from the local background. It is widely accepted in the target community that if this family of detectors could perform as desired, then these detectors are poised to find not only known targets as background anomalies, but also other potential targets that might not be known to the user.

However, a major disadvantage of using anomaly detectors, as discussed for instance in [1], is that they often produce an intolerable high number of meaningless detections (false alarms) for a given scene. The challenge of effective anomaly detection is further augmented when the sensor's line of sight is at the ground to ground (G2G) level (or approximately so), which corresponds to numerous Army applications. At this viewing perspective, the scales of objects in the scene are unknown, e.g., small targets will look large in the imagery for a small range. This scale uncertainty offers no guarantees that object samples in the imagery will be compared only to samples of their immediate surroundings, using, for instance, the well known inside-outside window approach [1] to supply samples to the particular detector. Scales of known targets can be reasonably estimated if the sensor's altitude is known; therefore, for typical surveillance operations where sensors are flown at known altitudes looking straight down (nadir look), the inside-outside window approach is suitable because one can place the outside window, WOUT, sufficiently apart from the inside window, WIN—although both are located concentrically—to ensure such that, at no location in the imagery, WIN shows a target and WOUT shows both background and target. A corrupted sample set shown in WOUT might be particularly damaging for detectors based on parametric distribution models for the clutter background (e.g., the industry standard RX anomaly detection is based on the multivariate Gaussian assumption, see [2]).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Adaptive Target-Scale-Invariant Hyperspectral Anomaly Detection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Armys Armament Research & Development Center (ARDEC) Picatinny, New Jersey, 07806**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

For robustness and adaptability purposes, we chose to use HS imagery and anomaly detection for the G2G viewing perspective. Recently, the use of HS imagery has also gained renewed attention in the target detection community. Its popularity over broadband imagery is due to the fact that these passive sensors simultaneously record images for hundreds of contiguous and narrowly spaced regions of the electromagnetic spectrum. Each image corresponds to the same ground scene, thus creating a cube of images that contain both spatial and spectral information about the objects and backgrounds in the scene. A host of different types of anomaly detectors and their performances in HSI are discussed in [1], [2], [3].

Our interest has been on a general idea where anomaly detectors perform a comparison between two observations by indirect means. The implementation of this idea has shown to preserve the number of meaningful anomaly detections and to significantly reduce the number of meaningless false positives, see [4].

In this paper, we present a compact test statistic for anomaly detection based on a principle of indirect comparison (PIC), see Section 2. We also present a random sampling approach that does not require secondary information about the targets (range), where a parallel process is introduced to mitigate the inclusion by chance of target samples into clutter background classes during the process of randomly sampling. The rationale here is that the likelihood of surpressing the presence of targets in all parallel processes is real however, relatively small.

The remaining of this paper is organized as follows: Section 2 formulates a PIC based anomaly detector, which in essence is an asymmetric variance test; Section 3 proposes a parallel random sampling approach, which makes possible for attaining a fully autonomous G2G anomaly detection without knowledge of object scales in the scene; Section 4 discusses the analyses performed on using results from Section 2 and Section 3; and Section 5 concludes this paper.

2. PIC-Based Anomaly Detector

The principle that leads to the design of a PIC based anomaly detector employs an indirect sample comparison to test the likelihood that local HS random samples belong to the same population. Let X and Y denote two random samples, and let $Z = X \cup Y$, where U denotes the union. We showed [4] that X can be indirectly compared to Y by comparing, instead, Z to Y (or to X). Using PIC, we propose next the asymmetric variance test (AVT) anomaly detector. Let random sequences x_0 and x_1 be observed according to the independent, identically distributed (iid) model

$$x_0 = (x_{01}, x_{02}, x_{03}, \ldots, x_{0n1}) \text{ iid} \sim g_0(x) \tag{1}$$

$$x_1 = (x_{11}, x_{12}, x_{13}, \ldots, x_{1n1}] \text{ iid} \sim g_1(x)$$

where, x_0 (test sample of size n_1) and x_1 (reference sample of size n_0) are independent, g_1 and g_0 are unknown, and

$$E(x_{0j}) = \mu_0 \; ; \; \mathrm{Var}(x_{0j}) = \sigma^2_0 < \infty \tag{2}$$
$$E(x_{1j}) = \mu_1 \; ; \; \mathrm{Var}(x_{1j}) = \sigma^2_1 < \infty \tag{3}$$
$$\mathrm{Var}[(x_{0j} - \mu_0)^2] = \zeta^2_0 < \infty \tag{4}$$

where, $E(\cdot)$ and $Var(\cdot)$ are expectation and variance operators. Now, consider the null hypothesis

$$H_0 : \; \sigma_0^2 = \tau \; (\tau > 0) \tag{5}$$

In (5), we would like to test the hypothesis that the variance from a reference sample is equal to an arbitrary positive value. At a first glance, the null hypothesis does not seem too effective as a discriminant feature, because τ can take any positive value, and additionally the variance, as a discriminant feature, does not account for the mean, which itself can be another discriminant feature. However, one can cleverly incorporate the indirect comparison approach discussed earlier to test (5), by designing in the process a rather effective anomaly detector. A solution follows.

Let the combined sample be represented by

$$t = (t_1, t_2, \ldots, t_n) = (x_{01}, \ldots, x_{0n1}, x_{11}, \ldots, x_{1n1}) \tag{6}$$

where, $n = n_1 + n_2$, and lets assume that its components have the same variance, i.e., $\mathrm{Var}(t_k) = \sigma^2_u$. The last assumption may not be satisfied for all t, but would certainly be satisfied when x_0 and x_1 are sampled from the same population, in which case one could set $\tau = \hat{\sigma}^2_u$ in (5), where $\hat{\sigma}^2_u$ estimates σ^2_u. Without the normality assumption in (1), deriving a test for the null hypothesis in (5) can be quite difficult. But as we anticipate a relatively large sample size in HSI, we shall rely on the *central limit theorem* (CLT) and the *weak law of large numbers* (WLLL), [5], to design the new detector. The set of parameters (μ_0, σ^2_0) then can be estimated by the following *consistent* estimators: $(\bar{x}_0, s_0^2)$, respectively, where

$$\bar{x} = \sum_{j=1}^{n_0} \frac{x_{0j}}{n_0}, \; s_0^2 = \sum_{j=1}^{n_0} \frac{\left(x_{0j} - \bar{x}_0\right)^2}{n_0 - 1} \tag{7}$$

Following (7), under general regularity conditions and using the denotations in (1), CLT ensures that the random variable z_1, below, converges in law to the standard Normal distribution $[N(0,1)]$, as the sample size n_0 increases, or

$$Z_1 = \sqrt{n_0}\,\frac{s_0^2 - \hat{\sigma}_0^2}{\sqrt{\xi_0^2}} \underset{n_0 \to \infty}{\rightarrow} N(0,1) \qquad (8)$$

To estimate ξ_0^2 using a consistent estimator $\left(\hat{\xi}_0^2\right)$, consider this rationale: Let ,

$$\vartheta_j = \left(x_{0j} - \mu_0\right)^2 \qquad (9)$$

and note that, based on (3) and (4),

$$E\left(\vartheta_j\right) = \sigma_0^2 : Var\left(\vartheta_j\right) = \xi_0^2 \qquad (10)$$

A consistent estimator of $Var\left(\vartheta_j\right)$ then would qualify for application in (8). An obvious estimator of $Var\left(\vartheta_j\right)$ is

$$\hat{V}_0 = \sum_{j=1}^{n_0} \frac{\left(\vartheta_j - \overline{\vartheta}\right)^2}{n_0 - 1} \qquad (11)$$

where $\overline{\vartheta}$ is the sample average using all ϑ_j's. Notice that $\hat{V}_0$ can be also expressed by the following decomposition,

$$\hat{V}_0 = \frac{n_0}{n_0 - 1} \sum_{j=1}^{n_0} \frac{\left(\vartheta_j - \sigma_0^2\right)^2 - \left(\overline{\vartheta} - \sigma_0^2\right)^2}{n_0} \qquad (12)$$

where the normalized summation term (which does not include $\overline{\vartheta}$) tends to ξ_0^2 in probability by the WLLN, and the term that includes $\overline{\vartheta}$ tends to *zero* in probability also by the WLLN. Therefore, $\hat{V}_0$ is a consistent estimator of ξ_0^2. In addition, using results from (7), notice that s_0^2 is also a consistent estimator of $E\left(\vartheta_j\right)$. We then propose the following consistent estimator of

$$\xi_0^2 = E\left\{ \left[\vartheta_j - E\left(\vartheta_j\right)\right]^2 \right\} \qquad (13)$$

to be

$$\hat{\xi}_0^2 = \sum_{j=1}^{n} \frac{\left[\left(x_{0j} - \bar{x}_0\right)^2 - s_0^2\right]^2}{n_0 - 1} \qquad (14)$$

Setting $\tau = \hat{\sigma}_u^2$ in (5), where

$$\hat{\sigma}_u^2 = \sum_{j=1}^{n} \frac{\left(t_j - \bar{t}\right)^2}{n - 1},\ \bar{t} = \sum_{j=1}^{n} \frac{t_j}{n} \qquad (15)$$

$$n = n_0 + n_1$$

if the null hypothesis in (5) is true, the following must also be true

$$Z_2 = \sqrt{n_0}\,\frac{s_0^2 - \hat{\sigma}_u^2}{\sqrt{\hat{\xi}_0^2}} \underset{n_0 \to \infty}{\rightarrow} N(0,1) \qquad (16)$$

Using properties of the family of chi square distributions (see for instance [5]), the following is also true under the null hypothesis:

$$Z_{AVT} = Z_2^2 = n_0 \frac{\left(s_0^2 - \hat{\sigma}_u^2\right)^2}{\hat{\xi}_0^2} \underset{n_0 \to \infty}{\rightarrow} \chi_1^2 \qquad (17)$$

where χ_1^2 is the chi-square probability density function (pdf) with *1* degree of freedom (dof). Testing hypothesis H_0 in (5) using (17) constitutes the AVT anomaly detector.

3. Random Sampling and Fusion Processes

For applications requiring G2G viewing perspective, full automation of the AVT algorithm can be only achieved if underlying difficulties (e.g., target scale uncertainties) can be handled effectively.

We propose in this section a random sampling technique and fusion of results to curb G2G perspective difficulties. The proposed random sampling technique and fusion of results can be simplified as follows: Suppose an N number of spectrum sets are randomly collected from the imagery, such that each set is tested against spectra from a window that moves across the imagery yielding N output surfaces from an anomaly detector. These N surfaces are fused by retaining the piecewise minimum at each pixel location in order to produce a single output surface (FMIN). In FMIN, blocks of data in the original imagery that are significantly different from the sampled N sets should be accentuated with respect to the spatial locations of the clutter background. In order to minimize the suppression probability of targets in the imagery by chance, a parallel process is introduced to repeat this sampling process M number of times and to finally yield a single decision surface by fusing (adding) these M FMIN surfaces.

An implicit assumption in this proposed approach is that the detector being used is quite effective suppressing both homogeneous regions and transitions of these regions, while capable to accentuate spectrally anomalous materials from background materials. If this assumption is satisfied, then as long as one out of the M FMIN surfaces has that anomalous object(s) accentuated, and M-1 FMIN surfaces have those same anomalous object(s) suppressed (due to the object[s] probability of

being selected by chance from the random sampling stage), the fusion of these M FMIN surfaces will show the accentuated anomalous object(s), as desired.

The number N of random samples collected per image needed depends on the level of background heterogeneity of the scene. Lesser heterogeneity requires a smaller N, and vice versa. In this paper, we study the overall approach by setting N = (10, 20, 50). Each random sample $\mathbf{R}_j$ ($j = 1, \cdots, N$) drawn from the image is defined by the transposed $[\cdot]$ vector

$$R_j = E(\mathbf{C}) = \left[\mu_{j1}, \cdots, \mu_{jb}\right]^t \tag{18}$$

where **C** has a size of *w* x *w* x *b*, *w* x *w* is the total number of pixels within the *jth* block of data, *b* is the number of bands, and μ_{ji} is the sample mean at the i^{th} band from an input size of *w* x *w*.

Fig. 1 shows the random locations of two experiments using N = 10 (white circles) and N = 50 (yellow circles), where representations of (18) can be assembled as

$$\mathbf{R} = \left[\mathbf{R}_1, \cdots, \mathbf{R}_N\right] \tag{19}$$

For a given N and a given image area *A*, the probability P of random locations landing at least *once* by chance at the target spatial area *a* within *A* can be modeled by the binomial distribution family as

$$\begin{aligned} P(n \geq 1) &= P(n=1) + P(n=2) + .. + P(n=N) \\ &= 1 - P(n=0) \end{aligned} \tag{20}$$

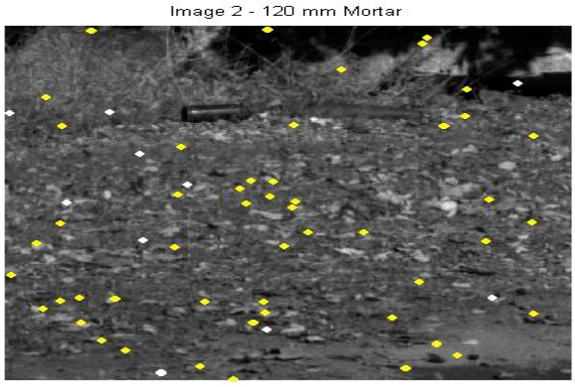

Figure 1. The SOC-700 HS sensor was used to collect this 640 x 640 pixel imagery by 120 bands, shown as the average of all bands. The background is formed by gravel, rocks, vegetation (tress and grass), and munitions (targets) lying in the ground—targets are shown at the top center. Random locations are shown for N = 10 (white circles) and N = 50 (yellow circles).

where

$$P(n) = \frac{N!}{n!(N-n)!} q^n (1-q)^{(N-n)}, \tag{21}$$

$n = 0, 1, \cdots, N$ and $q \in (0,1)$ is the ratio *a/A*.

Fig. 2 illustrates the relationship between *P* using (20) and N (represented in figure by RP) for two fixed values of *q* (*0.1* and *0.2*). In essence, Fig. 2 shows that the likelihood of drawing target samples by chance increases as N increases and for a reasonable N (e.g., N = 15), as *q* increases from *0.1* to *0.2*, *P* increases significantly, which is expected but undesired. It is desired to hold *P* at a relatively low value, while taking an adequate number N of clutter classes representing the entire background classes in the scene. To achieve this desired outcome out of this random sampling technique, we propose to employ a parallel sampling process, followed by a fusing step, as shown in Fig. 3.

Let S_m $(m = 1, \cdots, M)$ be an independent anomaly detection process, each using N independently collected random sample sets from an input cube $\mathbf{X}_{in}$. In the context of Section 2, S_m represents using model (1), the result in (17), and an additional fusing step to account for N independent inputs.

As shown in (20), each parallel process S_m in Fig. 3 has a probability *P* of blindly selecting target samples by chance during random sampling. If target samples are selected to represent a background class, the AVT detector will likely suppress that target class.

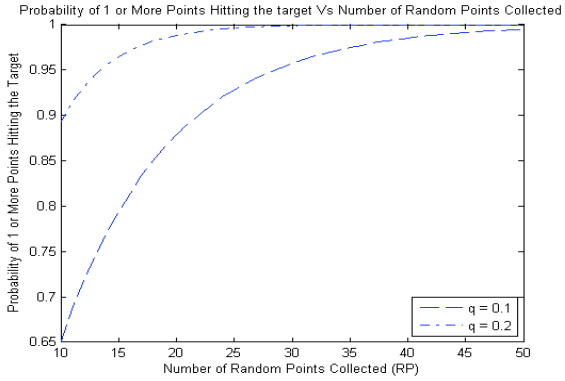

Figure 2. Relationship between the probability *P* of having at least one random point hitting on the target of size *q* and the chosen number of random points N. Parameter *q* is the ratio of the target size to the total image size.

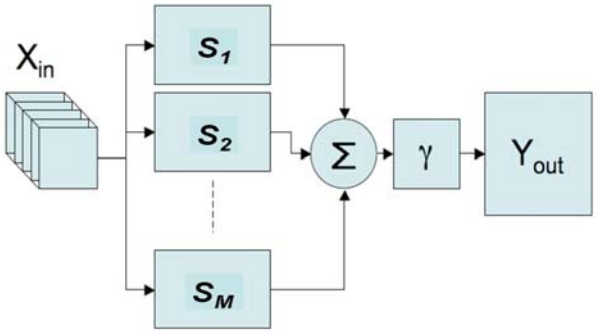

Figure 3 Parallel Process to mitigate target suppression by chance from final output result.

However, by fusing (adding) the detector's output from each $S_m \left(m = 1, \cdots, M\right)$, one would expect this probability to decrease. To account for this parallel processing, let P_T be the probability that all of the M processes happen to have the target selected by change during the processes' random sampling stage; this probability can be also modeled by the binomial distribution, or

$$P_T\left(m = M\right) = \frac{M!}{m!\left(M-m\right)!} P^m \left(1-P\right)^{\left(M-m\right)} = P^M \qquad (22)$$

where, $P = P\left(n \geq 1\right)$ in (20).

The benefit of using this parallel processing is shown in Fig. 4, where the relationship between P_T (vertical axis) and M (horizontal axis) is presented for two fixed values of P (0.65 and 0.90).

Having P_T decreasing relatively fast as M increases, the next step is to select an acceptable value of M from Fig. 4 and to combine the detectors' output results from the parallel process, as shown in Fig. 3, where a threshold γ is applied for a final decision. (This threshold can be attained empirically by experimenting offline with the detector and by estimating ROC curves, see Section 4.)

In summary, in reference to the random sampling step, by choosing an adequate value of N from Fig. 2, and by choosing an adequate value of M from Fig. 4, the user can compromise between having good representations of the background classes in the scene and maintaining a low probability of including target samples by chance as a background class.

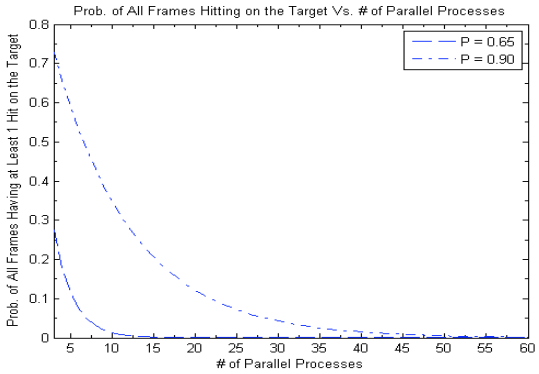

Figure 4. Probability that the random sampling stage of all the frames in the parallel process hit the target, given a target hit probability P (0.65 and 0.90) per frame.

4. Results

In order to study the behaviour of our general approach to G2G anomaly detection using real HS imagery, we designed a rather strict set of experiments using a large number of trials to test the HS imagery shown in Fig. 1 using the AVT anomaly detector. Results from these trials were used to estimate 95%-confidence-interval ROC curves representing different combinations of certain parameters. In addition to studying the behaviour of our approach, another important goal in this section is to show that the AVT detector's performance via ROC curves does not vary significantly if the random sampling stage of our approach is replaced by having a human in the loop selecting a priori a set of the most representative clutter background spectral samples in the given imaged scene.

In this section, P_d represents the proportion of the target that was detected by the AVT detector and P_{fa} represents the probability of false alarms (false positives), using a given threshold γ, as shown in Fig. 3.

As our major concern is to observe the behaviour of randomly sampling the scene as proposed, a large number of trials was required. The plots to follow were calculated by performing 2,730 trials for a parallel process of $M = 3$ and 3,000 trials for a parallel process of $M = 5$ for fixed values of $N = 10$, 20, and 50 randomly collected points. The ROC curves for each combination of parameter settings are presented as sample-average ROC curves out of the total number of trials. Each trial corresponds to employing the full parallel process discussed in Section 3 (see Fig. 3), using the HS cube shown in Fig. 1 as input and the AVT anomaly detector as discussed in Section 2 and Section 3.

Fig. 5 shows average ROC-curve comparison for settings M = 3 and N = 10, 20, 50 (random sampling) and N = 10 (manual sampling). Fig. 5 shows these curves to be virtually invariant to the different settings. Fig. 6 shows a zoomed in version of Fig. 5, by limiting the ranges of P_d and P_{fa}. The difference in performance for a given P_{fa}

between 10 points randomly or manually taken from the HS image, show an improvement of about 1% (Fig. 6). The performance between 10, 20, and 50 randomly collected points show an improvement from 10 to 20 and 50 of about 1 to 1.5%, but the performance between 20 and 50 seem to be similar. Depending on the value of *q*, as discussed in Section 3, an increase in performance of 1% may be desirable (as seen from 10 to 20 or 50 random points), but one needs to be aware that the probability of one of the randomly collected points hitting the target increases, as N increases, thus, a larger N may not be desirable.

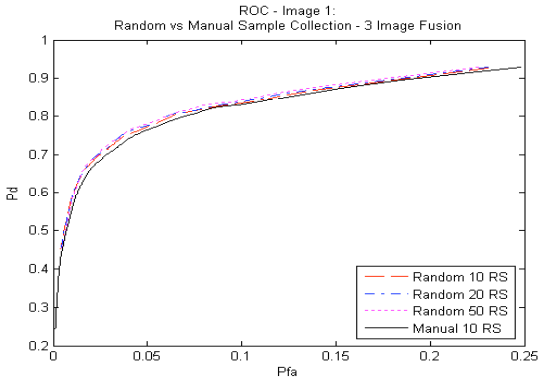

Figure 5. Average ROC curves using M = 3 and N = 10, 20, 50 (random sampling) and N = 10 (manual sampling) collected random points. Random points collected on the image surpassed manual collection by a small margin.

Fig. 7 illustrates 95% confidence intervals (using 2,730 trials) at two distinct points in the average ROC curves for the fixed settings: *M = 3* and *N = 10, 20, 50* (random points). These intervals are depicted in Fig. 7 as rectangular boxes in order to show how performance variability changes as a function of *N*. For instance, one can conclude from Fig. 7 that the confidence interval at the P_{fa} axis gets tighter as a function of increasing *N*. This tightness of the envelope is owed to a better statistical representation of the background using a larger N. Also, this variability was virtually the same for each parameter setting at two distinct points in the ROC curves.

Figures 8 and 9 reveal the impact of setting *M = 3* versus *M = 5* (parallel processes) and *N = 10, 50*. Although an increase in the number of fused processes (*N*) may be well suited to minimize the probability of eliminating the target of interest by the random sample collection, the impact of using either *M = 3* or *M = 5* parallel processes is minimum.

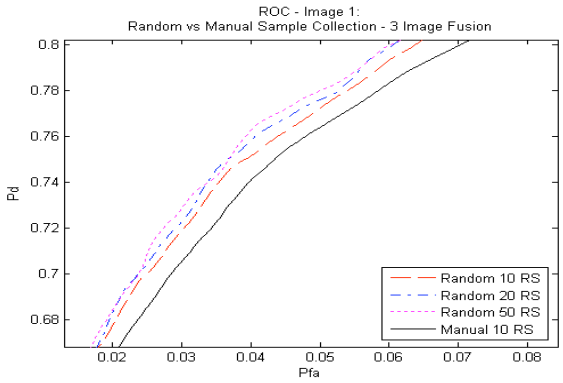

Figure 6. Zoomed in version of Fig. 5.

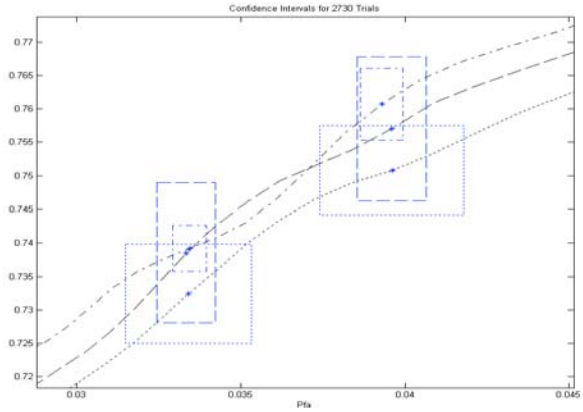

Figure 7. Confidence interval within Pd vs Pfa for 2730 trials for M = 3 (parallel processes). Number of random points are depicted as: 10 (..), 20 (--), and 50 (-.).

As seen in figure 9, using 5 over 3 parallel processes for both 10 and 50 random points produce very limited improvement in performance. Obviously, one cannot generalize the consistency of these results using a different anomaly detector and/or a different background/target scene. For future work, we intend to investigate the generalization of our findings in this paper by employing competing anomaly detection techniques and alternative background/target scenes.

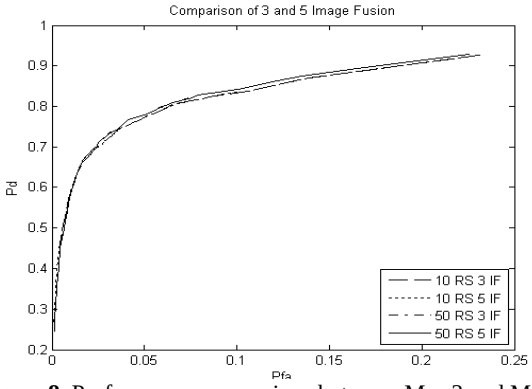

Figure 8. Performance comparison between M = 3 and M = 5

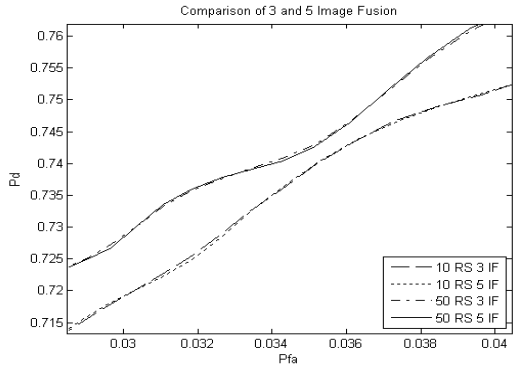

Figure 9. Zoomed in version of Fig. 8, see Pd and Pfa axes.

5. Conclusions

We proposed a fully autonomous parallel anomaly detection approach for challenging scenarios requiring ground to ground viewing perspectives.
At this viewing perspective, one cannot rely on approximations to object scales, which adds to the difficulties influencing the ultimate goal of attaining effective anomaly detection. Using HS imagery as input, the approach employs a random sampling stage to sample the imaged scene for background class representatives; a new anomaly detector, based on a principle of indirect comparison; and a parallel process, followed by fusion of results, to mitigate the likelihood of randomly selecting by chance samples of target spectra as background class representatives. The probability of randomly selecting target spectra was modeled by the binary distribution family, and helpful plots were presented to assist the user to address a compromise between autonomous random selection of an adequate number of spectra representing distinct background classes and holding down the probability of selecting target spectra by chance. To the best of our knowledge, the proposed approach is unique, and it should be of interest to the entire target community

References

[1] D. Manolakis, D. Marden, and G. Shaw, “Target detection algorithms for hyperspectral imaging applications,” *Lincoln Laboratory Journal*, 14(1), 2003.

[2] X. Yu, L. Hoff, I. Reed, A. Chen, L. Stotts, “Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach,” *IEEE Tran. Image Processing*, vol. 6, pp. 143-156, Jan. 1997.

[3] H. Kwon, and N. Nasrabadi, "Kernel Orthogonal Subspace Projection for Hyperspectral Target Classification", *IEEE Tran. Geoscience and Remote Sensing*, vol. 43, No. 12, pp 2952-2962, Dec. 2005.

[4] Anonymous

[5] E. L. Lehmann. *Testing Statistical Hypotheses*, New York: transferred to Chapman & Hall, 2nd Edition, pp 68, 74, Chapter 3, 1993.